1

Identity and Necessity

A problem which has arisen frequently in contemporary philosophy is: "How are *contingent* identity statements possible?" This question is phrased by analogy with the way Kant phrased his question "How are synthetic a priori judgments possible?" In both cases, it has usually been taken for granted in the one case by Kant that synthetic a priori judgments were possible, and in the other case in contemporary philosophical literature that contingent statements of identity are possible. I do not intend to deal with the Kantian question except to mention this analogy: After a rather thick book was written trying to answer the question how synthetic a priori judgments were possible, others came along later who claimed that the solution to the problem was that synthetic a priori judgments were, of course, impossible and that a book trying to show otherwise was written in vain. I will not discuss who was right on the possibility of synthetic a priori judgments. But in the case of contingent statements of identity, most philosophers have felt that the notion of a contingent identity statement ran into something like the following paradox. An argument like the following can be given against the possibility of contingent identity statements:[1]

First, the law of the substitutivity of identity says that, for any objects x and y, if x is identical to y, then if x has a certain property F, so does y:

[1] This paper was presented orally, without a written text, to a New York University lecture series on identity. The lecture was taped, and the present paper represents a transcription of these tapes, edited only slightly with no attempt to change the style of the original. If the reader imagines the sentences of this paper as being delivered, extemporaneously, with proper pauses and emphases, this may facilitate his comprehension. Nevertheless, there may still be passages which are hard to follow, and the time allotted necessitated a condensed presentation of the argument. (A longer version of some of these views, still rather compressed and still representing a transcript of oral remarks, appeared in *Semantics of Natural Language*, ed. by Donald Davidson and Gilbert Harman, [Dordrecht: D. Reidel, 1972]; it subsequently appeared as a monograph, titled *Naming and Necessity* [Cambridge: Harvard University Press, 1980].) Occasionally, reservations, amplifications and gratifications of my remarks had to be repressed, especially in the discussion of theoretical identification and the mind-body problem. The footnotes, which were added to the original, would have become even more unwieldy if this had not been done.

(1) $(x)(y)[(x = y) \supset (Fx \supset Fy)]$

On the other hand, every object surely is necessarily self-identical:

(2) $(x)\Box(x = x)$

But

(3) $(x)(y)(x = y) \supset [\Box(x = x) \supset \Box(x = y)]$

is a substitution instance of (1), the substitutivity law. From (2) and (3), we can conclude that, for every x and y, if x equals y, then, it is necessary that x equals y:

(4) $(x)(y)((x = y) \supset \Box(x = y))$

This is because the clause $\Box(x = x)$ of the conditional drops out because it is known to be true.

This is an argument which has been stated many times in recent philosophy. Its conclusion, however, has often been regarded as highly paradoxical. For example, David Wiggins, in his paper, "Identity-Statements," says,

> Now there undoubtedly exist contingent identity-statements. Let $a = b$ be one of them. From its simple truth and (5) [= (4) above] we can derive '$\Box(a = b)$'. But how then can there be any contingent identity-statements?[2]

He then says that five various reactions to this argument are possible, and rejects all of these reactions, and reacts himself. I do not want to discuss all the possible reactions to this statement, except to mention the second of those Wiggins rejects. This says,

> We might accept the result and plead that provided '*a*' and '*b*' are proper names nothing is amiss. The consequence of this is that no contingent identity-statements can be made by means of proper names.

And then he says that he is discontented with this solution and many other philosophers have been discontented with this solution, too, while still others have advocated it.

What makes the statement (4) seem surprising? It says, for any objects x and y, if x is y, then it is necessary that x is y. I have already mentioned that someone might object to this argument on the grounds that premise (2) is already false, that it is not the case that everything is necessarily self-identical. Well, for example, am I myself necessarily self-identical? Someone might argue that in some situations which we can imagine I would not even have existed and therefore the statement "Saul Kripke is Saul Kripke" would have been false or it would not be the case that I was self-identical. Perhaps, it would have been neither true nor false, in such a world, to say that Saul Kripke is self-identical.

[2] R. J. Butler, ed., *Analytical Philosophy, Second Series* (Oxford: Blackwell, 1965), p. 41.

Well, that may be so, but really it depends on one's philosophical view of a topic that I will not discuss, that is, what is to be said about truth values of statements mentioning objects that do not exist in the actual world or any given possible world or counterfactual situation. Let us interpret necessity here weakly. We can count statements as necessary if whenever the objects mentioned therein exist, the statement would be true. If we wished to be very careful about this, we would have to go into the question of existence as a predicate and ask if the statement can be reformulated in the form: For every x it is necessary that, if x exists, then x is self-identical. I will not go into this particular form of subtlety here because it is not going to be relevant to my main theme. Nor am I really going to consider formula (4). Anyone who believes formula (2) is, in my opinion, committed to formula (4). If x and y are the same things and we can talk about modal properties of an object at all, that is, in the usual parlance, we can speak of modality *de re* and an object *necessarily* having certain properties as such, then formula (1), I think, has to hold. Where x is any property at all, including a property involving modal operators, and if x and y are the same object and x had a certain property F, then y has to have the same property F. And this is so even if the property F is itself of the form of necessarily having some other property G, in particular that of necessarily being identical to a certain object. Well, I will not discuss the formula (4) itself because by itself it does not assert, of any particular true statement of identity, that it is necessary. It does not say anything about *statements* at all. It says for every *object* x and *object* y, if x and y are the same object, then it is necessary that x and y are the same object. And this, I think, if we think about it (anyway, if someone does not think so, I will not argue for it here), really amounts to something very little different from the statement (2). Since x, by definition of identity, is the only object identical with x, $(y)(y = x \supset Fy)$ seems to me to be little more than a garrulous way of saying 'Fx', and thus $(x)(y)(y = x \supset Fx)$ says the same as $(x)Fx$ no matter what 'F' is—in particular, even if 'F' stands for the property of necessary identity with x. So if x has this property (of necessary identity with x), trivially everything identical with x has it, as (4) asserts. But, from statement (4) one may apparently be able to deduce that various particular statements of identity must be necessary and this is then supposed to be a very paradoxical consequence.

Wiggins says, "Now there undoubtedly exist contingent identity-statements." One example of a contingent identity statement is the statement that the first Postmaster General of the United States is identical with the inventor of bifocals, or that both of these are identical with the man claimed by the *Saturday Evening Post* as its founder (*falsely* claimed, I gather, by the way). Now some such statements are plainly contingent. It plainly is a contingent fact that one and the same man both invented bifocals and took on the job of Postmaster General of the United States. How can we reconcile this with the truth of statement (4)? Well, that, too, is an issue I do not want to go into in detail except to be very dogmatic about it. It was I think settled quite well by Bertrand Russell in his

notion of the scope of a description. According to Russell, one can, for example, say with propriety that the author of *Hamlet* might not have written *Hamlet*, or even that the author of *Hamlet* might not have been the author of *Hamlet*. Now here, of course, we do not deny the necessity of the identity of an object with itself; but we say it is true concerning a certain man that he in fact was the unique person to have written *Hamlet* and secondly that the man, who in fact was the man who wrote *Hamlet*, might not have written *Hamlet*. In other words, if Shakespeare had decided not to write tragedies, he might not have written *Hamlet*. Under these circumstances, the man who in fact wrote *Hamlet* would not have written *Hamlet*. Russell brings this out by saying that in such a statement, the first occurrence of the description "the author of *Hamlet*" has large scope.[3] That is, we say "The author of *Hamlet* has the following property: that he might not have written *Hamlet*." We *do not* assert that the following statement might have been the case, namely that the author of *Hamlet* did not write *Hamlet*, for that is not true. That would be to say that it might have been the case that someone wrote *Hamlet* and yet did not write *Hamlet*, which would be a contradiction. Now, aside from the details of Russell's particular formulation of it, which depends on his theory of descriptions, this seems to be the distinction that any theory of descriptions has to make. For example, if someone were to meet the President of Harvard and take him to be a Teaching Fellow, he might say: "I took the President of Harvard for a Teaching Fellow." By this he does not mean that he took the proposition "The President of Harvard is a Teaching Fellow" to be true. He could have meant this, for example, had he believed that some sort of democratic system had gone so far at Harvard that the President of it decided to take on the task of being a Teaching Fellow. But that probably is not what he means. What he means instead, as Russell points out, is "Someone is President of Harvard and I took him to be a Teaching Fellow." In one of Russell's examples someone says, "I thought your yacht is much larger than it is." And the other man replies, "No, my yacht is not much larger than it is."

Provided that the notion of modality *de re*, and thus of quantifying into modal contexts, makes any sense at all, we have quite an adequate solution to the problem of avoiding paradoxes if we substitute descriptions for the universal quantifiers in (4) because the only consequence we will draw,[4] for example, in the bifocals case, is that there is a man who both happened to have invented bifocals

[3] The second occurrence of the description has small scope.

[4] In Russell's theory, $F(\iota xGx)$ follows from $(x)\ Fx$ and $(\exists!x)$ Gx, provided that the description in $F(\iota xGx)$ has the entire context for its scope (in Russell's 1905 terminology, has a 'primary occurrence'). Only then is $F(\iota xGx)$ 'about' the denotation of 'ιxGx'. Applying this rule to (4), we get the results indicated in the text. Notice that, in the ambiguous form $\Box(\iota xGx = \iota xHx)$, if one or both of the descriptions have 'primary occurrences' the formula does not assert the necessity of $\iota xGx = \iota xHx$; if both have secondary occurrences, it does. Thus in a language without explicit scope indicators, descriptions must be construed with the smallest possible scope—only then will $\sim A$ be the negation of A, $\Box A$ the necessitation of A, and the like. See Chapter 8.

and happened to have been the first Postmaster General of the United States, and is necessarily self-identical. There is an object *x* such that *x* invented bifocals, and as a matter of contingent fact an object *y*, such that *y* is the first Postmaster General of the United States, and finally, it is necessary, that *x* is *y*. What are *x* and *y* here? Here, *x* and *y* are both Benjamin Franklin, and it can certainly be necessary that Benjamin Franklin is identical with himself. So, there is no problem in the case of descriptions if we accept Russell's notion of scope.[5] And I just dogmatically want to drop that question here and go on to the question about names which Wiggins raises. And Wiggins says he might accept the result and plead that, provided *a* and *b* are proper names, nothing is amiss. And then he rejects this.

Now what is the special problem about proper names? At least if one is not familiar with the philosophical literature about this matter, one naively feels something like the following about proper names. First, if someone says "Cicero was an orator," then he uses the name 'Cicero' in that statement simply to pick out a certain object and then to ascribe a certain property to the object, namely, in this case, he ascribes to a certain man the property of having been an orator. If someone else uses another name, such as, say, 'Tully', he is still speaking about the same man. One ascribes the same property, if one says "Tully is an orator," to the same man. So to speak, the fact, or state of affairs, represented by the statement is the same whether one says "Cicero is an orator" or one says "Tully is an orator." It would, therefore, seem that the function of names is *simply* to refer, and not to describe the objects so named by such properties as "being the inventor of bifocals" or "being the first Postmaster General." It would seem that Leibniz' law and the law (1) should not only hold in the universally quantified form, but also in the form "if $a = b$ and Fa, then Fb," wherever 'a' and 'b' stand in

[5] An earlier distinction with the same purpose was, of course, the medieval one of *de dicto-de re*. That Russell's distinction of scope eliminates modal paradoxes has been pointed out by many logicians, especially Smullyan.

So as to avoid misunderstanding, let me emphasize that I am of course not asserting that Russell's notion of scope solves Quine's problem of 'essentialism'; what it does show, especially in conjunction with modern model-theoretic approaches to modal logic, is that quantified modal logic need not deny the truth of all instances of $(x)(y)(x = y \cdot \supset \cdot Fx \supset Fy)$, nor of all instances of '$(x)(Gx \supset Ga)$' (where 'a' is to be replaced by a nonvacuous definite description whose scope is all of 'Ga'), in order to avoid making it a necessary truth that one and the same man invented bifocals and headed the original Postal Department. Russell's contextual definition of description need not be adopted in order to ensure these results; but other logical theories, Fregean or other, which take descriptions as primitive must somehow express the same logical facts. Frege showed that a simple, non-iterated context containing a definite description with small scope, which cannot be interpreted as being 'about' the denotation of the description, can be interpreted as about its 'sense'. Some logicians have been interested in the question of the conditions under which, in an intensional context, a description with small scope is equivalent to the same one with large scope. One of the virtues of a Russellian treatment of descriptions in modal logic is that the answer (roughly that the description be a 'rigid designator' in the sense of this lecture) then often follows from the other postulates for quantified modal logic: no special postulates are needed, as in Hintikka's treatment. Even if descriptions are taken as primitive, special postulation of when scope is irrelevant can often be deduced from more basic axioms. (See Chapter 11.)

place of names and '*F*' stands in place of a predicate expressing a genuine property of the object:

$(a = b \cdot Fa) \supset Fb$

We can run the same argument through again to obtain the conclusion where '*a*' and '*b*' replace any names, "If *a* = *b*, then necessarily *a* = *b*." And so, we could venture this conclusion: that whenever '*a*' and '*b*' are proper names, if *a* is *b*, that it is necessary that *a* is *b*. Identity statements between proper names have to be necessary if they are going to be true at all. This view in fact has been advocated, for example, by Ruth Barcan Marcus in a paper of hers on the philosophical interpretation of modal logic.[6] According to this view, whenever, for example, someone makes a correct statement of identity between two names, such as, for example, that Cicero is Tully, his statement has to be necessary if it is true. But such a conclusion *seems* plainly to be false. (I, like other philosophers, have a habit of understatement in which "it seems plainly false" means "it is plainly false." Actually, I think the view is true, though not quite in the form defended by Mrs. Marcus.) At any rate, it seems plainly false. One example was given by Professor Quine in his reply to Professor Marcus at the symposium: "I think I see trouble anyway in the contrast between proper names and descriptions as Professor Marcus draws it. The paradigm of the assigning of proper names is tagging. We may tag the planet Venus some fine evening with the proper name 'Hesperus'. We may tag the same planet again someday before sun rise with the proper name 'Phosphorus'." (Quine thinks that something like that actually was done once.) "When, at last, we discover that we have tagged the same planet twice, our discovery is empirical, and not because the proper names were descriptions." According to what we are told, the planet Venus seen in the morning was originally thought to be a star and was called "the Morning Star," or (to get rid of any question of using a description) was called 'Phosphorus'. One and the same planet, when seen in the evening, was thought to be another star, the Evening Star, and was called "Hesperus." Later on, astronomers discovered that Phosphorus and Hesperus were one and the same. Surely no amount of a priori ratiocination on their part could conceivably have made it possible for them to deduce that Phosphorus is Hesperus. In fact, given the information they had, it might have turned out the other way. Therefore, it is argued, the statement 'Hesperus is Phosphorus' has to be an ordinary contingent, empirical truth, one which might have come out otherwise, and so the view that true identity statements between names are necessary has to be false. Another example which Quine gives in *Word and Object* is taken from Professor Schrödinger, the famous pioneer of quantum mechanics: A certain mountain can be seen from

[6] "Modalities and Intensional Languages," *Boston Studies in the Philosophy of Science*, Vol. 1 (New York: Humanities Press, 1963), pp. 71ff. See also the "Comments" by Quine and the ensuing discussion.

both Tibet and Nepal. When seen from one direction it was called 'Gaurisanker'; when seen from another direction, it was called 'Everest'; and then, later on, the empirical discovery was made that Gaurisanker *is* Everest. (Quine further says that he gathers the example is actually geographically incorrect. I guess one should not rely on physicists for geographical information.)

Of course, one possible reaction to this argument is to deny that names like 'Cicero', 'Tully', 'Gaurisanker', and 'Everest' really are proper names. Look, someone might say (someone has said it: his name was 'Bertrand Russell'), just because statements like "Hesperus is Phosphorus" and "Gaurisanker is Everest" are contingent, we can see that the names in question are not really purely referential. You are not, in Mrs. Marcus' phrase, just 'tagging' an object; you are actually describing it. What does the contingent fact that Hesperus is Phosphorus amount to? Well, it amounts to the fact that *the* star in a certain portion of the sky in the evening is *the* star in a certain portion of the sky in the morning. Similarly, the contingent fact that Guarisanker is Everest amounts to the fact that the mountain viewed from such and such an angle in Nepal is the mountain viewed from such and such another angle in Tibet. Therefore, such names as 'Hesperus' and 'Phosphorus' can only be abbreviations for descriptions. The term 'Phosphorus' *has* to mean "the star seen . . . ," or (let us be cautious because it actually turned out not to be a star), "the *heavenly body* seen from such and such a position at such and such a time in the morning," and the name 'Hesperus' has to mean "the heavenly body seen in such and such a position at such and such a time in the evening." So, Russell concludes, if we want to reserve the term "name" for things which really just name an object without describing it, the only real proper names we can have are names of our own immediate sense data, objects of our own 'immediate acquaintance'. The only such names which occur in language are demonstratives like "this" and "that." And it is easy to see that this requirement of necessity of identity, understood as exempting identities between names from all imaginable doubt, can indeed be guaranteed only for demonstrative names of immediate sense data; for only in such cases can an identity statement between two different names have a general immunity from Cartesian doubt. There are some other things Russell has sometimes allowed as objects of acquaintance, such as one's self; we need not go into details here. Other philosophers (for example, Mrs. Marcus in her reply, at least in the verbal discussion as I remember it—I do not know if this got into print, so perhaps this should not be 'tagged' on her[7]) have said, "If names are really just tags, genuine tags, then a good dictionary should be able to tell us that they are names of the same object." You have an object *a* and an object *b* with names 'John' and 'Joe'. Then, according to Mrs. Marcus, a dictionary should be able to tell you whether or not 'John' and 'Joe' are names of the same object. Of course, I do not

[7] It should. See her remark on p. 115, *Boston Studies in the Philosophy of Science*, Vol. 1, in the discussion following the papers.

know what ideal dictionaries should do, but ordinary proper names do not seem to satisfy this requirement. You certainly *can*, in the case of ordinary proper names, make quite empirical discoveries that, let's say, Hesperus is Phosphorus, though we thought otherwise. We can be in doubt as to whether Gaurisanker is Everest or Cicero is in fact Tully. Even now, we could conceivably discover that we were wrong in supposing that Hesperus was Phosphorus. Maybe the astronomers made an error. So it seems that this view is wrong and that if by a name we do not mean some artificial notion of names such as Russell's, but a proper name in the ordinary sense, then there can be contingent identity statements using proper names, and the view to the contrary seems plainly wrong.

In recent philosophy a large number of other identity statements have been emphasized as examples of contingent identity statements, different, perhaps, from either of the types I have mentioned before. One of them is, for example, the statement "Heat is the motion of molecules." First, science is supposed to have discovered this. Empirical scientists in their investigations have been supposed to discover (and, I suppose, they did) that the external phenomenon which we call "heat" is, in fact, molecular agitation. Another example of such a discovery is that water is H_2O, and yet other examples are that gold is the element with such and such an atomic number, that light is a stream of photons, and so on. These are all in some sense of "identity statement" identity statements. Second, it is thought, they are plainly contingent identity statements, just because they were scientific discoveries. After all, heat might have turned out not to have been the motion of molecules. There were other alternative theories of heat proposed, for example, the caloric theory of heat. If these theories of heat had been correct, then heat would not have been the motion of molecules, but instead, some substance suffusing the hot object, called "caloric." And it was a matter of course of science and not of any logical necessity that the one theory turned out to be correct and the other theory turned out to be incorrect.

So, here again, we have, apparently, another plain example of a contingent identity statement. This has been supposed to be a very important example because of its connection with the mind-body problem. There have been many philosophers who have wanted to be materialists, and to be materialists in a particular form, which is known today as "the identity theory." According to this theory, a certain mental state, such as a person's being in pain, is identical with a certain state of his brain (or, perhaps, of his entire body, according to some theorists), at any rate, a certain material or neural state of his brain or body. And so, according to this theory, my being in pain at this instant, if I were, would be identical with my body's being or my brain's being in a certain state. Others have objected that this cannot be because, after all, we can imagine my pain existing even if the state of the body did not. We can perhaps imagine my not being embodied at all and still being in pain, or, conversely, we could imagine my body existing and being in the very same state even if there were no pain. In fact, conceivably, it could be in this state even though there were no mind 'back of it',

so to speak, at all. The usual reply has been to concede that all of these things might have been the case, but to argue that these are irrelevant to the question of the identity of the mental state and the physical state. This identity, it is said, is just another contingent scientific identification, similar to the identification of heat with molecular motion, or water with H_2O. Just as we can imagine heat without any molecular motion, so we can imagine a mental state without any corresponding brain state. But, just as the first fact is not damaging to the identification of heat and the motion of molecules, so the second fact is not at all damaging to the identification of a mental state with the corresponding brain state. And so, many recent philosophers have held it to be very important for our theoretical understanding of the mind-body problem that there can be contingent identity statements of this form.

To state finally what *I* think, as opposed to what seems to be the case, or what others think, I think that in both cases, the case of names and the case of the theoretical identifications, the identity statements are necessary and not contingent. That is to say, they are necessary if *true*; of course, false identity statements are not necessary. How can one possibly defend such a view? Perhaps I lack a complete answer to this question, even though I am convinced that the view is true. But to begin an answer, let me make some distinctions that I want to use. The first is between a *rigid* and a *nonrigid designator*. What do these terms mean? As an example of a nonrigid designator, I can give an expression such as 'the inventor of bifocals'. Let us suppose it was Benjamin Franklin who invented bifocals, and so the expression, 'the inventor of bifocals', designates or refers to a certain man, namely, Benjamin Franklin. However, we can easily imagine that the world could have been different, that under different circumstances someone else would have come upon this invention before Benjamin Franklin did, and in that case, *he* would have been the inventor of bifocals. So, in this sense, the expression 'the inventor of bifocals' is nonrigid: Under certain circumstances one man would have been the inventor of bifocals; under other circumstances, another man would have. In contrast, consider the expression 'the square root of 25'. Independently of the empirical facts, we can give an arithmetical proof that the square root of 25 is in fact the number 5, and because we have proved this mathematically, what we have proved is necessary. If we think of numbers as entities at all, and let us suppose, at least for the purpose of this lecture, that we do, then the expression 'the square root of 25' necessarily designates a certain number, namely 5. Such an expression I call 'a *rigid* designator'. Some philosophers think that anyone who even uses the notions of rigid or nonrigid designator has already shown that he has fallen into a certain confusion or has not paid attention to certain facts. What do I mean by 'rigid designator'? I mean a term that designates the same object in all possible worlds. To get rid of one confusion which certainly is not mine, I do not use "might have designated a different object" to refer to the fact that language might have been used differently. For example, the expression 'the inventor of bifocals' might have been used by

inhabitants of this planet always to refer to the man who corrupted Hadleyburg. This would have been the case, if, first, the people on this planet had not spoken English, but some other language, which phonetically overlapped with English; and if, second, in that language the expression 'the inventor of bifocals' meant the 'man who corrupted Hadleyburg'. Then it would refer, of course, in their language, to whoever in fact corrupted Hadleyburg in this counterfactual situation. That is not what I mean. What I mean by saying that a description might have referred to something different, I mean that in *our* language as *we* use it in describing a counterfactual situation, there might have been a different object satisfying the descriptive conditions *we* give for reference. So, for example, we use the phrase 'the inventor of bifocals', when we are talking about another possible world or a counterfactual situation, to refer to whoever in that counterfactual situation would have invented bifocals, not to the person whom people *in* that counterfactual situation would have called 'the inventor of bifocals'. *They* might have spoken a different language which phonetically overlapped with English in which 'the inventor of bifocals' is used in some other way. I am *not* concerned with that question here. For that matter, they might have been deaf and dumb, or there might have been no people at all. (There still could have been an inventor of bifocals even if there were no people—God, or Satan, will do.)

Second, in talking about the notion of a rigid designator, I do not mean to imply that the object referred to has to exist in all possible worlds, that is, that it has to necessarily exist. Some things, perhaps mathematical entities such as the positive integers, if they exist at all, necessarily exist. Some people have held that God both exists and necessarily exists; others, that He contingently exists; others, that He contingently fails to exist; and others, that He necessarily fails to exist:[8] all four options have been tried. But at any rate, when I use the notion of rigid designator, I do not imply that the object referred to necessarily exists. All I mean is that in any possible world where the object in question *does* exist, in any situation where the object *would* exist, we use the designator in question to designate that object. In a situation where the object does not exist, then we should say that the designator has no referent and that the object in question so designated does not exist.

As I said, many philosophers would find the very notion of rigid designator objectionable per se. And the objection that people make may be stated as follows: Look, you're talking about situations which are counterfactual, that is to say, you're talking about other possible worlds. Now these worlds are completely disjoint, after all, from the actual world which is not just another possible world; it is the actual world. So, before you talk about, let us say, such an object as Richard Nixon in another possible world at all, you have to say which object in this other possible world would *be* Richard Nixon. Let us talk about a

[8] If there is no deity, and especially if the nonexistence of a deity is *necessary*, it is dubious that we can use "He" to refer to a deity. The use in the text must be taken to be non-literal.

situation in which, as *you* would say, Richard Nixon would have been a member of SDS. Certainly the member of SDS you are talking about is someone very different in many of his properties from Nixon. Before we even can say whether this man would have been Richard Nixon or not, we have to set up criteria of identity across possible worlds. Here are these other possible worlds. There are all kinds of objects in them with different properties from those of any actual object. Some of them resemble Nixon in some ways, some of them resemble Nixon in other ways. Well, which of these objects is Nixon? One has to give a criterion of identity. And this shows how the very notion of rigid designator runs in a circle. Suppose we designate a certain number as the number of planets. Then, if that is our favorite way, so to speak, of designating this number, then in any other possible worlds we will have to identify whatever number is the number of planets with the number 9, which in the actual world is the number of planets. So, it is argued by various philosophers, for example, implicitly by Quine, and explicitly by many others in his wake, we cannot really ask whether a designator is rigid or nonrigid because we first need a criterion of identity across possible worlds. An extreme view has even been held that, since possible worlds are so disjoint from our own, we cannot really say that any object in them is the *same* as an object existing now but only that there are some objects which resemble things in the actual world, more or less. We, therefore, should not really speak of what would have been true of Nixon in another possible world but, only of what 'counterparts' (the term which David Lewis uses[9]) of Nixon there would have been. Some people in other possible worlds have dogs whom they call 'Checkers'. Others favor the ABM but do not have any dog called Checkers. There are various people who resemble Nixon more or less, but none of them can really be said to be Nixon; they are only *counterparts* of Nixon, and you choose which one is the best counterpart by noting which resembles Nixon the most closely, according to your favorite criteria. Such views are widespread, both among the defenders of quantified modal logic and among its detractors.

All of this talk seems to me to have taken the metaphor of possible worlds much too seriously in some way. It is as if a 'possible world' were like a foreign country, or distant planet way out there. It is as if we see dimly through a telescope various actors on this distant planet. Actually David Lewis' view seems the most reasonable if one takes this picture literally. No one far away on another planet can be strictly identical with someone here. But, even if we have some marvelous methods of transportation to take one and the same person from planet to planet, we really need some epistemological criteria of identity to be able to say whether someone on this distant planet is the same person as someone here.

All of this seems to me to be a totally misguided way of looking at things. What it amounts to is the view that counterfactual situations have to be described

[9] David K. Lewis, "Counterpart Theory and Quantified Modal Logic," *Journal of Philosophy*, LXV (1968), 113ff.

purely qualitatively. So, we cannot say, for example, "If Nixon had only given a sufficient bribe to Senator X, he would have gotten Carswell through" because that refers to certain people, Nixon and Carswell, and talks about what things would be true of them in a counterfactual situation. We must say instead "If a man who has a hairline like such and such, and holds such and such political opinions had given a bribe to a man who was a senator and had such and such other qualities, then a man who was a judge in the South and had many other qualities resembling Carswell would have been confirmed." In other words, we must describe counterfactual situations purely qualitatively and then ask the question, "Given that the situation contains people or things with such and such qualities, which of these people is (or is a counterpart of) Nixon, which is Carswell, and so on?" This seems to me to be wrong. Who is to prevent us from saying "Nixon might have gotten Carswell through had he done certain things"? We are speaking of *Nixon* and asking what, in certain counterfactual situations, would have been true of *him*. We can say that if Nixon had done such and such, he would have lost the election to Humphrey. Those I am opposing would argue, "Yes, but how do you find out if the man you are talking about is in fact Nixon?" It would indeed be very hard to find out, if you were looking at the whole situation through a telescope, but that is not what we are doing here. Possible worlds are not something to which an epistemological question like this applies. And if the phrase 'possible worlds' is what makes anyone think some such question applies, he should just *drop* this phrase and use some other expression, say 'counterfactual situation,' which might be less misleading. If we say "If Nixon had bribed such and such a Senator, Nixon would have gotten Carswell through," what is *given* in the very description of that situation is that it is a situation in which we are speaking of Nixon, and of Carswell, and of such and such a Senator. And there seems to be no less objection to *stipulating* that we are speaking of certain *people* than there can be objection to stipulating that we are speaking of certain *qualities*. Advocates of the other view take speaking of certain qualities as unobjectionable. They do not say, "How do we know that this quality (in another possible world) is that of redness?" But they do find speaking of certain *people* objectionable. But I see no more reason to object in the one case than in the other. I think it really comes from the idea of possible worlds as existing out there, but very far off, viewable only through a special telescope. Even more objectionable is the view of David Lewis. According to Lewis, when we say "Under certain circumstances Nixon would have gotten Carswell through," we really mean "Some man, other than Nixon but closely resembling him, would have gotten some judge, other than Carswell but closely resembling him, through." Maybe that is so, that some man closely resembling Nixon could have gotten some man closely resembling Carswell through. But *that* would not comfort either Nixon or Carswell, nor would it make Nixon kick himself and say "*I* should have done such and such to get Carswell through." The question is whether under certain circumstances Nixon *himself* could have

gotten *Carswell* through. And I think the objection is simply based on a misguided picture.

Instead, we can perfectly well talk about rigid and nonrigid designators. Moreover, we have a simple, intuitive test for them. We can say, for example, that the number of planets might have been a different number from the number it in fact is. For example, there might have been only seven planets. We can say that the inventor of bifocals might have been someone other than the man who *in fact* invented bifocals.[10] We cannot say, though, that the square root of 81 might have been a different number from the number it in fact is, for that number just has to be 9. If we apply this intuitive test to proper names, such as for example 'Richard Nixon', they would seem intuitively to come out to be rigid designators. First, when we talk even about the counterfactual situation in which we suppose Nixon to have done different things, we assume we are still talking about Nixon himself. We say, "If Nixon had bribed a certain Senator, he would have gotten Carswell through," and we assume that by 'Nixon' and 'Carswell' we are still referring to the very same people as in the actual world. And it seems that we cannot say "Nixon might have been a different man from the man he in fact was," unless, of course, we mean it metaphorically: He might have been a different *sort* of person (if you believe in free will and that people are not inherently corrupt). You might think the statement true in that sense, but Nixon could not have been in the other literal sense a different person from the person he, in fact, is, even though the thirty-seventh President of the United States might have been Humphrey. So the phrase "the thirty-seventh President" is nonrigid, but 'Nixon', it would seem, is rigid.

Let me make another distinction before I go back to the question of identity statements. This distinction is very fundamental and also hard to see through. In recent discussion, many philosophers who have debated the meaningfulness of various categories of truths, have regarded them as identical. Some of those who identify them are vociferous defenders of them, and others, such as Quine, say they are all identically meaningless. But usually they're not distinguished. These

[10] Some philosophers think that definite descriptions, in English, are ambiguous, that sometimes 'the inventor of bifocals' rigidly designates the man who in fact invented bifocals. I am tentatively inclined to reject this view, construed as a thesis about English (as opposed to a possible hypothetical language), but I will not argue the question here.

What I do wish to note is that, contrary to some opinions, this alleged ambiguity cannot replace the Russellian notion of the scope of a description. Consider the sentence, "The number of planets might have been necessarily even." This sentence plainly can be read so as to express a truth; had there been eight planets, the number of planets would have been necessarily even. Yet without scope distinctions, both a 'referential' (rigid) and a non-rigid reading of the description will make the statement false. (Since the number of planets is nine, the rigid reading amounts to the falsity that nine might have been necessarily even.)

The 'rigid' reading is equivalent to the Russellian primary occurrence; the non-rigid, to innermost scope—some, following Donnellan, perhaps loosely, have called this reading the 'attributive' use. The possibility of intermediate scopes is then ignored. In the present instance, the intended reading of $\Diamond\Box$ (the number of planets is even) makes the scope of the description $\Box$ (the number of planets is even), neither the largest nor the smallest possible. (See also Chapter 8, section 2a.)

are categories such as 'analytic', 'necessary', 'a priori', and sometimes even 'certain'. I will not talk about all of these but only about the notions of a prioricity and necessity. Very often these are held to be synonyms. (Many philosophers probably should not be described as holding them to be synonyms; they simply *use* them interchangeably.) I wish to distinguish them. What do we mean by calling a statement *necessary*? We simply mean that the statement in question, first, is true, and, second, that it could not have been otherwise. When we say that something is *contingently* true, we mean that, though it is in fact the case, it could have been the case that things would have been otherwise. If we wish to assign this distinction to a branch of philosophy, we should assign it to metaphysics. To the contrary, there is the notion of an *a priori truth*. An a priori truth is supposed to be one which can be *known* to be true independently of all experience. Notice that this does not in and of itself say anything about all possible worlds, unless this is put into the definition. All that it says is that it can be known to be true of the actual world, independently of all experience. It may, by some philosophical argument, follow from our knowing, independently of experience, that something is true of the actual world, that it has to be known to be true also of all possible worlds. But if this is to be established, it requires some philosophical argument to establish it. Now, *this* notion, if we were to assign it to a branch of philosophy, belongs, not to metaphysics, but to epistemology. It has to do with the way we can know certain things to be in fact true. Now, it may be the case, of course, that anything which is necessary is something which *can* be known a priori. (Notice, by the way, the notion a priori truth as thus defined has in it *another* modality: it *can* be known independently of all experience. It is a little complicated because there is a double modality here.) I will not have time to explore these notions in full detail here, but one thing we can see from the outset is that these two notions are by no means trivially the same. If they are coextensive, it takes some philosophical argument to establish it. As stated, they belong to different domains of philosophy. One of them has something to do with *knowledge*, of what can be known in certain ways about the *actual* world. The other one has to do with *metaphysics*, how the world *could* have been; given that it is the way it is, could it have been otherwise, in certain ways? Now I hold, as a matter of fact, that neither class of statements is contained in the other. But, all we need to talk about here is this: Is everything that is necessary knowable a priori or known a priori? Consider the following example: the Goldbach conjecture. This says that every even number is the sum of two primes. It is a mathematical statement and if it is true at all, it has to be necessary. Certainly, one could not say that though in fact every even number is the sum of two primes, there could have been some extra number which was even and not the sum of two primes. What would that mean? On the other hand, the answer to the question whether every even number *is* in fact the sum of two primes is unknown, and we have no method at present for deciding. So we certainly do not know, a priori or even a posteriori, that every even number is the sum of two primes. (Well, perhaps we

have some evidence in that no counterexample has been found.) But we certainly do not know a priori anyway, that every even number is, in fact, the sum of two primes. But, of course, the definition just says "*can* be known independently of experience," and someone might say that if it is true, we *could* know it independently of experience. It is hard to see exactly what this claim means. It might be so. One thing it might mean is that if it were true we could *prove* it. This claim is certainly wrong if it is generally applied to mathematical statements and we have to work within some fixed system. This is what Gödel proved. And even if we mean an 'intuitive proof in general' it might just be the case (at least, this view is as clear and as probable as the contrary) that though the statement is true, there is just no way the human mind could ever prove it. Of course, one way an *infinite* mind might be able to prove it is by looking through each natural number one by one and checking. In this sense, of course, it can, perhaps, be known a priori, but only by an infinite mind, and then this gets into other complicated questions. I do not want to discuss questions about the conceivability of performing an infinite number of acts like looking through each number one by one. A vast philosophical literature has been written on this: Some have declared it is logically impossible; others that it is logically possible; and some do not know. The main point is that it is not trivial that just because such a statement is necessary it can be known a priori. Some considerable clarification is required before we decide that it can be so known. And so this shows that even if everything necessary is a priori in some sense, it should not be taken as a trivial matter of definition. It is a substantive philosophical thesis which requires some work.

Another example that one might give relates to the problem of essentialism. Here is a lectern. A question which has often been raised in philosophy is: What are its essential properties? What properties, aside from trivial ones like self-identity, are such that this object has to have them if it exists at all,[11] are such that if an object did not have it, it would not be this object?[12] For example, being

[11] This definition is the usual formulation of the notion of essential property, but an exception must be made for existence itself: on the definition given, existence would be trivially essential. We should regard existence as essential to an object only if the object necessarily exists. Perhaps there are other recherché properties, involving existence, for which the definition is similarly objectionable. (I thank Michael Slote for this observation.)

[12] The two clauses of the sentence footnoted give equivalent definitions of the notion of essential property, since $\Box ((\exists x)(x=a) \supset Fa)$ is equivalent to $\Box(x)(\sim Fx \supset x=a)$. The second formulation, however, has served as a powerful seducer in favor of theories of 'identification across possible worlds'. For it suggests that we consider 'an object *b* in another possible world' and test whether it is identifiable with *a* by asking whether it lacks any of the essential properties of *a*. Let me therefore emphasize that, although an essential property is (trivially) a property without which an object cannot be *a*, it by no means follows that the essential, purely qualitative properties of *a* jointly form a sufficient condition for being *a*, nor that *any* purely qualitative conditions are sufficient for an object to be *a*. Further, even if necessary and sufficient qualitative conditions for an object to be Nixon may exist, there would still be little justification for the demand for a purely qualitative description of all counterfactual situations. We can ask whether Nixon might have been a Democrat without engaging in these subtleties.

made of wood, and not of ice, might be an essential property of this lectern. Let us just take the weaker statement that it is not made of ice. That will establish it as strongly as we need it, perhaps as dramatically. Supposing this lectern is in fact made of wood, could this very lectern have been made from the very beginning of its existence from ice, say frozen from water in the Thames? One has a considerable feeling that it could *not*, though in fact one certainly could have made a lectern of water from the Thames, frozen it into ice by some process, and put it right there in place of this thing. If one had done so, one would have made, of course, a *different* object. It would not have been *this very lectern*, and so one would not have a case in which this very lectern here was made of ice, or was made from water from the Thames. The question of whether it could afterward, say in a minute from now, turn into ice is something else. So, it would seem, if an example like this is correct—and this is what advocates of essentialism have held—that this lectern could not have been made of ice, that is in any counterfactual situation of which we would say that this lectern existed at all, we would have to say also that it was not made from water from the Thames frozen into ice. Some have rejected, of course, any such notion of essential property as meaningless. Usually, it is because (and I think this is what Quine, for example, would say) they have held that it depends on the notion of identity across possible worlds, and that this is itself meaningless. Since I have rejected this view already, I will not deal with it again. We can talk about *this very object*, and whether it could have had certain properties which it does not in fact have. For example, it could have been in another room from the room it in fact is in, even at this very time, but it could not have been made from the very beginning from water frozen into ice.

If the essentialist view is correct, it can only be correct if we sharply distinguish between the notions of a posteriori and a priori truth on the one hand, and contingent and necessary truth on the other hand, for although the statement that this table, if it exists at all, was not made of ice, is necessary, it certainly is not something that we know a priori. What we know is that first, lecterns usually are not made of ice, they are usually made of wood. This looks like wood. It does not feel cold and it probably would if it were made of ice. Therefore, I conclude, probably this is not made of ice. Here my entire judgment is a posteriori. I could find out that an ingenious trick has been played upon me and that, in fact, this lectern is made of ice; but what I am saying is, given that it is in fact not made of ice, in fact is made of wood, one cannot imagine that under certain circumstances it could have been made of ice. So we have to say that though we cannot know a priori whether this table was made of ice or not, given that it is not made of ice, it is *necessarily* not made of ice. In other words, if *P* is the statement that the lectern is not made of ice, one knows by a priori philosophical analysis, some conditional of the form "if *P*, then necessarily *P*." If the table is not made of ice, it is necessarily not made of ice. On the other hand, then, we know by empirical investigation that *P*, the antecedent of the conditional, is true—that this table is not made of ice. We can conclude by *modus ponens*:

$$P \supset \Box P$$

$$\frac{P}{\Box P}$$

The conclusion—'$\Box P$'—is that it is necessary that the table not be made of ice, and this conclusion is known a posteriori, since one of the premises on which it is based is a posteriori. So, the notion of essential properties can be maintained only by distinguishing between the notions of a priori and necessary truth, and I do maintain it.

Let us return to the question of identities. Concerning the statement 'Hesperus is Phosphorus' or the statement 'Cicero is Tully', one can find all of these out by empirical investigation, and we might turn out to be wrong in our empirical beliefs. So, it is usually argued, such statements must therefore be contingent. Some have embraced the other side of the coin and have held "Because of this argument about necessity, identity statements between names have to be knowable a priori, so, only a very special category of names, possibly, really works as names; the other things are bogus names, disguised descriptions, or something of the sort. However, a certain very narrow class of statements of identity are known a priori, and these are the ones which contain the genuine names." If one accepts the distinctions that I have made, one need not jump to either conclusion. One can hold that certain statements of identity between names, though often known a posteriori, and maybe not knowable a priori, are in fact necessary, if true. So, we have some room to hold this. But, of course, to have some room to hold it does not mean that we should hold it. So let us see what the evidence is. First, recall the remark that I made that proper names seem to be rigid designators, as when we use the name 'Nixon' to talk about a certain man, even in counterfactual situations. If we say, "If Nixon had not written the letter to Saxbe, maybe he would have gotten Carswell through," we are in this statement talking about Nixon, Saxbe, and Carswell, the very same men as in the actual world, and what would have happened to them under certain counterfactual circumstances. If names are rigid designators, then there can be no question about identities being necessary, because '*a*' and '*b*' will be rigid designators of a certain man or thing *x*. Then even in every possible world, *a* and *b* will both refer to this same object *x*, and to no other, and so there will be no situation in which *a* might not have been *b*. That would have, to be a situation in which the object which we are also now calling '*x*' would not have been identical with itself. Then one could not possibly have a situation in which Cicero would not have been Tully or Hesperus would not have been Phosphorus.[13]

[13] I thus agree with Quine, that "Hesperus is Phosphorus" is (or can be) an empirical discovery; with Marcus, that it is necessary. Both Quine and Marcus, according to the present standpoint, err in identifying the epistemological and the metaphysical issues.

Aside from the identification of necessity with a priority, what has made people feel the other way? There are two things which have made people feel the other way.[14] Some people tend to regard identity statements as metalinguistic statements, to identify the statement "Hesperus is Phosphorus" with the metalinguistic statement, "'Hesperus' and 'Phosphorus' are names of the same heavenly body." And that, of course, might have been false. We might have used the terms 'Hesperus' and 'Phosphorus' as names of *two* different heavenly bodies. But, of course, this has nothing to do with the necessity of identity. In the same sense "2 + 2 = 4" might have been false. The phrases "2 + 2" and "4" might have been used to refer to two different numbers. One can imagine a language, for example, in which "+", "2", and "=" were used in the standard way, but "4" was used as the name of, say, the square root of minus 1, as we should call it, "*i*." Then "2 + 2 = 4" would be false, for 2 plus 2 is not equal to the square root of minus 1. But this is not what we want. We do not want just to say that a certain statement which we in fact use to express something true could have expressed something false. We want to use the statement in *our* way and see if it could have been false. Let us do this. What is the idea people have? They say, "Look, Hesperus might not have been Phosphorus. Here a certain planet was seen in the morning, and it was seen in the evening; and it just turned out later on as a matter of empirical fact that they were one and the same planet. If things had turned out otherwise, they would have been two different planets, or two different heavenly bodies, so how can you say that such a statement is necessary?"

Now there are two things that such people can mean. First, they can mean that we do not know a priori whether Hesperus is Phosphorus. This I have already conceded. Second, they may mean that they can actually imagine circumstances that they would call circumstances in which Hesperus would not have been Phosphorus. Let us think what would be such a circumstance, using these terms here as *names* of a planet. For example, it could have been the case that Venus did indeed rise in the morning in exactly the position in which we saw it, but that on the other hand, in the position which is in fact occupied by Venus in the evening, Venus was not there, and Mars took its place. This is all counterfactual because in fact Venus is there. Now one can also imagine that in this counterfactual other

[14] The two confusions alleged, especially the second, are both related to the confusion of the metaphysical question of the necessity of "Hesperus is Phosphorus" with the epistemological question of its a prioricity. For if Hesperus is identified by its position in the sky in the evening, and Phosphorus by its position in the morning, an investigator may well know, in advance of empirical research, that Hesperus is Phosphorus if and only if one and the same body occupies position *x* in the evening and position *y* in the morning. The a priori material equivalence of the two statements, however, does not imply their strict (necessary) equivalence. (The same remarks apply to the case of heat and molecular motion below.) Similar remarks apply to some extent to the relationship between "Hesperus is Phosphorus" and "'Hesperus' and 'Phosphorus' name the same thing." A confusion that also operates is, of course, the confusion between what *we* say of a counterfactual situation and how people *in* that situation would have described it; this confusion, too, is probably related to the confusion between a prioricity and necessity.

possible world, the earth would have been inhabited by people and that they should have used the names 'Phosphorus' for Venus in the morning and 'Hesperus' for Mars in the evening. Now, this is all very good, but would it be a situation in which Hesperus was not Phosphorus? Of course, it is a situation in which people would have been able to *say*, truly, "Hesperus is not Phosphorus"; but we are supposed to describe things in our language, not in theirs. So let us describe it in our language. Well, how could it actually happen that Venus would not be in that position in the evening? For example, let us say that there is some comet that comes around every evening and yanks things over a little bit. (That would be a very simple scientific way of imagining it: not really too simple—that is very hard to imagine actually.) It just happens to come around every evening, and things get yanked over a bit. Mars gets yanked over to the very position where Venus is, then the comet yanks things back to their normal position in the morning. Thinking of this planet which we now call 'Phosphorus', what should we say? Well, we can say that the comet passes it and yanks Phosphorus over so that it is not in the position normally occupied by Phosphorus in the evening. If we do say this, and really use 'Phosphorus' as the name of a planet, then we have to say that, under such circumstances, Phosphorus in the evening would not be in the position where we, in fact, saw it; or alternatively, Hesperus in the evening would not be in the position in which we, in fact, saw it. We might say that under such circumstances, we would not have called Hesperus 'Hesperus' because Hesperus would have been in a different position. But that still would not make Phosphorus different from Hesperus; but what would then be the case instead is that Hesperus would have been in a different position from the position it in fact is and, perhaps, not in such a position that people would have called it 'Hesperus'. But that would not be a situation in which Phosphorus would not have been Hersperus.

Let us take another example which may be clearer. Suppose someone uses 'Tully' to refer to the Roman orator who denounced Cataline and uses the name 'Cicero' to refer to the man whose works he had to study in third-year Latin in high school. Of course, he may not know in advance that the very same man who denounced Cataline wrote these works, and that is a contingent statement. But the fact that this statement is contingent should not make us think that the statement that Cicero is Tully, if it is true, and it is in fact true, is contingent. Suppose, for example, that Cicero actually did denounce Cataline, but thought that this political achievement was so great that he should not bother writing any literary works. Would we say that these would be circumstances under which he would not have been Cicero? It seems to me that the answer is no, that instead we would say that, under such circumstances, Cicero would not have written any literary works. It is not a necessary property of Cicero—the way the shadow follows the man—that he should have written certain works; we can easily imagine a situation in which Shakespeare would not have written the works of Shakespeare, or one in which Cicero would not have written the works of Cicero.

What may be the case is that we *fix the reference* of the term 'Cicero' by use of some descriptive phrase, such as 'the author of these works'. But once we have this reference fixed, we then use the name 'Cicero' *rigidly* to designate the man who in fact we have identified by his authorship of these works. We do not use it to designate whoever would have written these works in place of Cicero, if someone else wrote them. It might have been the case that the man who wrote these works was not the man who denounced Cataline. Cassius might have written these works. But we would not then say that Cicero would have been Cassius, unless we were speaking in a very loose and metaphorical way. We would say that Cicero, whom we may have identified and come to know by his works, would not have written them, and that someone else, say Cassius, would have written them in his place.

Such examples are not grounds for thinking that identity statements are contingent. To take them as such grounds is to misconstrue the relation between a *name* and a *description used to fix its reference*, to take them to be *synonyms*. Even if we fix the reference of such a name as 'Cicero' as the man who wrote such and such works, in speaking of counterfactual situations, when we speak of Cicero, we do not then speak of whoever in such counterfactual situations *would* have written such and such works, but rather of Cicero, whom we have identified by the contingent property that he is the man who in fact, that is, in the actual world, wrote certain works.[15]

I hope this is reasonably clear in a brief compass. Now, actually I have been presupposing something I do not really believe to be, in general, true. Let us suppose that we do fix the reference of a name by a description. Even if we do so, we do not then make the name *synonymous* with the description, but instead we use the name *rigidly* to refer to the object so named, even in talking about counterfactual situations where the thing named would not satisfy the description in question. Now, this is what I think in fact is true for those cases of naming where the reference is fixed by description. But, in fact, I also think, contrary to most recent theorists, that the reference of names is rarely or almost never fixed by means of description. And by this I do not just mean what Searle says: "It's not

[15] If someone protests, regarding the lectern, that it *could* after all have *turned out* to have been made of ice, and therefore could have been made of ice, I would reply that what he really means is that *a lectern* could have looked just like this one, and have been placed in the same position as this one, and yet have been made of ice. In short, I could have been in the *same epistemological situation* in relation to *a lectern made of ice* as I actually am in relation to *this* lectern. In the main text, I have argued that the same reply should be given to protests that Hesperus could have turned out to be other than Phosphorus, or Cicero other than Tully. Here, then, the notion of 'counterpart' comes into its own. For it is not this table, but an epistemic 'counterpart', which was hewn from ice; not Hesperus-Phosphorus-Venus, but two distinct counterparts thereof, in two of the roles Venus actually plays (that of Evening Star and Morning Star), which are different. Precisely because of this fact, it is not *this table* which could have been made of ice. Statements about the modal properties of *this table* never refer to counterparts. However, if someone confuses the epistemological and the metaphysical problems, he will be well on the way to the counterpart theory Lewis and others have advocated.

a single description, but rather a cluster, a family of properties which fixes the reference." I mean that properties in this sense are not used *at all.* But I do not have the time to go into this here. So, let us suppose that at least one half of prevailing views about naming is true, that the reference is fixed by descriptions. Even were that true, the name would not be synonymous with the description, but would be used to *name* an object which we pick out by the contingent fact that it satisfies a certain description. And so, even though we can imagine a case where the man who wrote these works would not have been the man who denounced Cataline, we should not say that that would be a case in which Cicero would not have been Tully. We should say that it is a case in which Cicero did not write these works, but rather that Cassius did. And the identity of Cicero and Tully still holds.

Let me turn to the case of heat and the motion of molecules. Here surely is a case that is contingent identity! Recent philosophy has emphasized this again and again. So, if it is a case of contingent identity, then let us imagine under what circumstances it would be false. Now, concerning this statement I hold that the circumstances philosophers apparently have in mind as circumstances under which it would have been false are not in fact such circumstances. First, of course, it is argued that "Heat is the motion of molecules" is an a posteriori judgment; scientific investigation might have turned out otherwise. As I said before, this shows nothing against the view that it is necessary—at least if I am right. But here, surely, people had very specific circumstances in mind under which, so they thought, the judgment that heat is the motion of molecules would have been false. What were these circumstances? One can distill them out of the fact that we found out empirically that heat is the motion of molecules. How was this? What did we find out first when we found out that heat is the motion of molecules? There is a certain external phenomenon which we can sense by the sense of touch, and it produces a sensation which we call "the sensation of heat." We then discover that the external phenomenon which produces this sensation, which we sense, by means of our sense of touch, is in fact that of molecular agitation in the thing that we touch, a very high degree of molecular agitation. So, it might be thought, to imagine a situation in which heat would not have been the motion of molecules, we need only imagine a situation in which we would have had the very same sensation and it would have been produced by something other than the motion of molecules. Similarly, if we wanted to imagine a situation in which light was not a stream of photons, we could imagine a situation in which we were sensitive to something else in exactly the same way, producing what we call visual experiences, though not through a stream of photons. To make the case stronger, or to look at another side of the coin, we could also consider a situation in which we *are* concerned with the motion of molecules but in which such motion does not give us the sensation of heat. And it might also have happened that we, or, at least, the creatures inhabiting this planet, might have been so constituted that, let us say, an increase in the motion

of molecules did not give us this sensation but that, on the contrary, a slowing down of the molecules did give us the very same sensation. This would be a situation, so it might be thought, in which heat would not be the motion of molecules, or, more precisely, in which temperature would not be mean molecular kinetic energy.

But I think it would not be so. Let us think about the situation again. First, let us think about it in the actual world. Imagine right now the world invaded by a number of Martians, who do indeed get the very sensation that we call "the sensation of heat" when they feel some ice which has slow molecular motion, and who do not get a sensation of heat—in fact, maybe just the reverse—when they put their hand near a fire which causes a lot of molecular agitation. Would we say, "Ah, this casts some doubt on heat being the motion of molecules, because there are these other people who don't get the same sensation"? Obviously not, and no one would think so. We would say instead that the Martians somehow feel the very sensation we get when we feel heat when they feel cold and that they do not get a sensation of heat when they feel heat. But now let us think of a counterfactual situation.[16] Suppose the earth had from the very beginning been inhabited by such creatures. First, imagine it inhabited by no creatures at all: then there is no one to feel any sensations of heat. But we would not say that under such circumstances it would necessarily be the case that heat did not exist; we would say that heat might have existed, for example, if there were fires that heated up the air.

Let us suppose the laws of physics were not very different: Fires do heat up the air. Then there would have been heat even though there were no creatures around to feel it. Now let us suppose evolution takes place, and life is created, and there are some creatures around. But they are not like us, they are more like the Martians. Now would we say that heat has suddenly turned to cold, because of the way the creatures of this planet sense it? No, I think we should describe this situation as a situation in which, though the creatures on this planet got our sensation of heat, they did not get it when they were exposed to heat. They got it when they were exposed to cold. And that is something we can surely well imagine. We can imagine it just as we can imagine our planet being invaded by creatures of this sort. Think of it in two steps. First there is a stage where there are no creatures at all, and one can certainly imagine the planet still having both heat and cold, though no one is around to sense it. Then the planet comes through an evolutionary process to be peopled with beings of different neural

[16] Isn't the situation I just described also counterfactual? At least it may well be, if such Martians never in fact invade. Strictly speaking, the distinction I wish to draw compares how we *would* speak *in* a (possibly counterfactual) situation, *if* it obtained, and how we *do* speak *of* a counterfactual situation, knowing that it does not obtain—i.e., the distinction between the language we would have used in a situation and the language we *do* use to describe it. (Consider the description: "Suppose we all spoke German." This description is in English.) The former case can be made vivid by imagining the counterfactual situation to be actual.

structure from ourselves. Then these creatures could be such that they were insensitive to heat; they did not feel it in the way we do; but on the other hand, they felt cold in much the same way that we feel heat. But still, heat would be heat, and cold would be cold. And particularly, then, this goes in no way against saying that in this counterfactual situation heat would still *be* the molecular motion, *be* that which is produced by fires, and so on, just as it would have been if there had been no creatures on the planet at all. Similarly, we could imagine that the planet was inhabited by creatures who got visual sensations when there were sound waves in the air. We should not therefore say, "Under such circumstances, sound would have been light." Instead we should say, "The planet was inhabited by creatures who were in some sense visually sensitive to sound, and maybe even visually sensitive to light." If this is correct, it can still be and will still be a necessary truth that heat is the motion of molecules and that light is a stream of photons.

To state the view succinctly: we use both the terms 'heat' and 'the motion of molecules' as rigid designators for a certain external phenomenon. Since heat is in fact the motion of molecules, and the designators are rigid, by the argument I have given here, it is going to be *necessary* that heat is the motion of molecules. What gives us the illusion of contingency is the fact we have identified the heat by the contingent fact that there happen to be creatures on this planet—(namely, ourselves) who are sensitive to it in a certain way, that is, who are sensitive to the motion of molecules or to heat—these are one and the same thing. And this is contingent. So we use the description, 'that which causes such and such sensations, or that which we sense in such and such a way', to identify heat. But in using this fact we use a contingent property of heat, just as we use the contingent property of Cicero as having written such and such works to identify him. We then use the terms 'heat' in the one case and 'Cicero' in the other *rigidly* to designate the objects for which they stand. And of course the term 'the motion of molecules' is rigid; it always stands for the motion of molecules, never for any other phenomenon. So, as Bishop Butler said, "everything is what it is and not another thing." Therefore, "Heat is the motion of molecules" will be necessary, not contingent, and one only has the *illusion* of contingency in the way one could have the illusion of contingency in thinking that this table might have been made of ice. We might think one could imagine it, but if we try, we can see on reflection that what we are really imagining is just there being another lectern in this very position here which was in fact made of ice. The fact that we may identify this lectern by being the object we see and touch in such and such a position is something else.

Now how does this relate to the problem of mind and body? It is usually held that this is a contingent identity statement just like "Heat is the motion of molecules." That cannot be. It cannot be a contingent identity statement just like "Heat is the motion of molecules" because, if I am right, "Heat is the motion of molecules" is not a contingent identity statement. Let us look at this statement. For example, "My being in pain at such and such a time is my being in such and

such a brain state at such and such a time," or, "Pain in general is such and such a neural (brain) state."

This is held to be contingent on the following grounds. First, we can imagine the brain state existing though there is no pain at all. It is only a scientific fact that whenever we are in a certain brain state we have a pain. Second, one might imagine a creature being in pain, but not being in any specified brain state at all, maybe not having a brain at all. People even think, at least prima facie, though they may be wrong, that they can imagine totally disembodied creatures, at any rate certainly not creatures with bodies anything like our own. So it seems that we can imagine definite circumstances under which this relationship would have been false. Now, if these circumstances are circumstances, notice that we cannot deal with them simply by saying that this is just an illusion, something we can apparently imagine, but in fact cannot in the way we thought erroneously that we could imagine a situation in which heat was not the motion of molecules. Because although we can say that we pick out heat contingently by the contingent property that it affects us in such and such a way, we cannot similarly say that we pick out pain contingently by the fact that it affects us in such and such a way. On such a picture there would be the brain state, and we pick it out by the contingent fact that it affects us as pain. Now that might be true of the brain state, but it cannot be true of the pain. The experience itself has to be *this experience*, and I cannot say that it is a contingent property of the pain I now have that it is a pain.[17] In fact, it would seem that both the terms, 'my pain' and 'my being in such and such a brain state' are, first of all, both rigid designators. That is, whenever anything is such and such a pain, it is essentially that very object, namely, such and such a pain, and wherever anything is such and such a brain

[17] The most popular identity theories advocated today explicitly fail to satisfy this simple requirement. For these theories usually hold that a mental state is a brain state, and that what makes the brain state into a mental state is its 'causal role', the fact that it tends to produce certain behavior (as intentions produce actions, or pain, pain behavior) and to be produced by certain stimuli (e.g., pain, by pinpricks). If the relations between the brain state and its causes and effects are regarded as contingent, then *being such-and-such-a-mental-state* is a contingent property of the brain state. Let X be a pain. The causal-role identity theorist holds (1) that X is a brain state, (2) that the fact that X is a pain is to be analyzed (roughly) as the fact that X is produced by certain stimuli and produces certain behavior. The fact mentioned in (2) is, of course, regarded as contingent; the brain state X might well exist and not tend to produce the appropriate behavior in the absence of other conditions. Thus (1) and (2) assert that a certain pain X might have existed, yet not have been a pain. This seems to me self-evidently absurd. Imagine any pain: is it possible that *it itself* could have existed, yet not have been a pain?

If $X = Y$, then X and Y share all properties, including modal properties. If X is a pain and Y the corresponding brain state, then *being a pain* is an essential property of X, and *being a brain state* is an essential property of Y. If the correspondence relation is, in fact, identity, then it must be *necessary* of Y that it corresponds to a pain, and *necessary* of X that it correspond to a brain state, indeed to this particular brain state, Y. Both assertions seem false; it *seems* clearly possible that X should have existed without the corresponding brain state; or that the brain state should have existed without being felt as pain. Identity theorists cannot, contrary to their almost universal present practice, accept these intuitions; they must deny them, and explain them away. This is none too easy a thing to do.

state, it is essentially that very object, namely, such and such a brain state. So both of these are rigid designators. One cannot say this pain might have been something else, some other state. These are both rigid designators.

Second, the way we would think of picking them out—namely, the pain by its being an experience of a certain sort, and the brain state by its being the state of a certain material object, being of such and such molecular configuration—both of these pick out their objects essentially and not accidentally, that is, they pick them out by essential properties. Whenever the molecules *are* in this configuration, we *do* have such and such a brain state. Whenever you feel *this*, you do have a pain. So it seems that the identity theorist is in some trouble, for, since we have two rigid designators, the identity statement in question is necessary. Because they pick out their objects essentially, we cannot say the case where you seem to imagine the identity statement false is really an illusion like the illusion one gets in the case of heat and molecular motion, because that illusion depended on the fact that we pick out heat by a certain contingent property. So there is very little room to maneuver; perhaps none.[18] The identity theorist, who holds that pain is the brain state, also has to hold that it necessarily is the brain state. He therefore cannot concede, but has to deny, that there would have been situations under which one would have had pain but not the corresponding brain state. Now usually in arguments on the identity theory, this is very far from being denied. In fact, it is conceded from the outset by the materialist as well as by his opponent. He says, "Of course, it *could* have been the case that we had pains without the brain states. It is a contingent identity." But that cannot be. He has to hold that we are under some illusion in thinking that we can imagine that there could have been pains without brain states. And the only model I can think of for what the illusion might be, or at least the model given by the analogy the materialists

[18] A brief restatement of the argument may be helpful here. If "pain" and "C-fiber stimulation" are rigid designators of phenomena, one who identifies them must regard the identity as necessary. How can this necessity be reconciled with the apparent fact that C-fiber stimulation might have turned out not to be correlated with pain at all? We might try to reply by analogy to the case of heat and molecular motion; the latter identity, too, is necessary, yet someone may believe that, before scientific investigation showed otherwise, molecular motion might have turned out not to be heat. The reply is, of course, that what really is possible is that people (or some rational or sentient beings) could have been in the *same epistemic situation* as we actually are, and identify *a phenomenon* in the same way we identify heat, namely, by feeling it by the sensation we call "the sensation of heat," without the phenomenon being molecular motion. Further, the beings might not have been sensitive to molecular motion (i.e., to heat) by any neural mechanism whatsoever. It is impossible to explain the apparent possibility of C-fiber stimulations not having been pain in the same way. Here, too, we would have to suppose that we could have been in the same epistemological situation, and identify something in the same way we identify pain, without its corresponding to C-fiber stimulation. But the way we identify pain is by feeling it, and if a C-fiber stimulation could have occurred without our feeling any pain, then the C-fiber stimulation would have occurred without there *being* any pain, contrary to the necessity of the identity. The trouble is that although 'heat' is a rigid designator, heat is picked out by the contingent property of its being felt in a certain way; pain, on the other hand, is picked out by an essential (indeed necessary and sufficient) property. For a sensation to be *felt* as pain is for it to *be* pain.

themselves suggest, namely, heat and molecular motion, simply does not work in this case. So the materialist is up against a very stiff challenge. He has to show that these things we think we can see to be possible are in fact not possible. He has to show that these things which we can imagine are not in fact things we can imagine. And that requires some very different philosophical argument from the sort which has been given in the case of heat and molecular motion. And it would have to be a deeper and subtler argument than I can fathom and subtler than has ever appeared in any materialist literature that I have read. So the conclusion of this investigation would be that the analytical tools we are using go against the identity thesis and so go against the general thesis that mental states are just physical states.[19]

The next topic would be my own solution to the mind-body problem, but that I do not have.

[19] All arguments against the identity theory which rely on the necessity of identity, or on the notion of essential property, are, of course, inspired by Descartes' argument for his dualism. The earlier arguments which superficially were rebutted by the analogies of heat and molecular motion, and the bifocals inventor who was also Postmaster General, had such an inspiration: and so does my argument here. R. Albritton and M. Slote have informed me that they independently have attempted to give essentialist arguments against the identity theory, and probably others have done so as well.

The simplest Cartesian argument can perhaps be restated as follows: Let '*A*' be a *name* (rigid designator) of Descartes' body. Then Descartes argues that since he could exist even if *A* did not, $\Diamond$(Descartes $\neq$ A), hence Descartes $\neq$ *A*. Those who have accused him of a modal fallacy have forgotten that '*A*' is rigid. His argument is valid, and his conclusion is correct, provided its (perhaps dubitable) premise is accepted. On the other hand, provided that Descartes is regarded as having ceased to exist upon his death, "Descartes $\neq$ *A*" can be established without the use of a modal argument; for if so, no doubt *A* survived Descartes when *A* was a corpse. Thus *A* had a property (existing at a certain time) which Descartes did not. The same argument can establish that a statue is not the hunk of stone, or the congery of molecules, of which it is composed. Mere non-identity, then, may be a weak conclusion. (See D. Wiggins, *Philosophical Review*, LXXVII [1968], 90ff.) The Cartesian modal argument, however, surely can be deployed to maintain relevant stronger conclusions as well.